EVIDENCE FOR THE POLYPHYLY OF THE SCROPHULARIACEAE BASED ON CHLOROPLAST *rbc*L AND *ndh*F SEQUENCES[1]

Richard G. Olmstead[2] *and*
Patrick A. Reeves[3]

ABSTRACT

Data from the chloroplast genes *rbc*L and *ndh*F, totaling more than 3500 base pairs of DNA sequence, were used to examine the monophyly of the Scrophulariaceae, including several groups that have been suggested to belong to or to be derived from the Scrophulariaceae. Thirty-two taxa representing the Lamiales s.l. and outgroups were sampled and each of the sets of gene sequences was analyzed separately and in combination. Results indicate that two distinct clades composed of elements of the traditional Scrophulariaceae exist and that a monophyletic Scrophulariaceae, even one liberally circumscribed to include several small families, cannot be supported by these data. One group, designated "scroph I," includes *Verbascum*, *Celsia*, *Selago*, *Scrophularia*, *Buddleja*, and *Nicodemia*. A second group, "scroph II," includes *Antirrhinum*, *Digitalis*, *Veronica*, and the Plantaginaceae, Callitrichaceae, and Hippuridaceae. *Schlegelia* and *Paulownia*, often assigned either to the Scrophulariaceae or Bignoniaceae, do not appear with either family.

The Scrophulariaceae are a well-known family to temperate botanists and one for which a "gestalt" recognition serves the scientist well. Members of the family are generally recognized by their usually herbaceous habit, their typically bilaterally symmetric, tubular flowers, and their many-seeded capsular fruits. Taxonomists learn the exceptions to the "usual," such as the woody *Paulownia* trees and the nearly actinomorphic flowers of *Verbascum*, and learn the characters that identify related groups: the winged seeds lacking endosperm of the Bignoniaceae, the achlorophyllous parasites of the Orobanchaceae, the wind-pollinated Plantaginaceae, the reduced ovule number in the Labiatae and Verbenaceae, and the inflorescence bracts and explosive capsules of the Acanthaceae. However, the characters by which we recognize plants as scrophs cannot be labeled shared derived traits, or synapomorphies, leaving open the possibility that the family is not monophyletic.

The taxonomic history of the Scrophulariaceae reflects the problem of describing a natural group. Most recent angiosperm classifications (e.g., Cronquist, 1981; Takhtajan, 1987; Thorne, 1992) have recognized a suprafamilial group centered around the Scrophulariaceae and containing approximately 12 (e.g., Scrophulariales of Cronquist, 1981) to 15 (e.g., Scrophulariinae of Thorne, 1992) families. Some of the more clearly defined families in the group are suggested to be connected to the Scrophulariaceae by intermediate genera (e.g., *Nelsonia* connecting with the Acanthaceae or *Paulownia* and *Schlegelia* connecting with the Bignoniaceae), whereas other families (e.g., Orobanchaceae, Plantaginaceae, Lentibulariaceae) "may logically be considered to be specialized derivatives of the Scrophulariaceae" (Cronquist, 1981). The Scrophulariaceae, as circumscribed in the above-mentioned treatments, usually have the largest number of species of any family in the group (e.g., 3000, Thorne, 1992). Most recent treatments of the Scrophulariaceae and related families are derived from the 19th century treatment of Bentham & Hooker (Bentham, 1876), who maintained separate families for the Lentibulariaceae, Orobanchaceae, Plantaginaceae, and Selaginaceae. Wettstein (1895) included the Selaginaceae as a tribe in the Scrophulariaceae. Hallier (1903), in recognition of the apparently derived nature of the Lentibulariaceae, Orobanchaceae, Plantaginaceae,

[1] This research has been supported by NSF grant BSR-9107827 to RGO and a University of Colorado, Undergraduate Research Opportunities Program grant to PAR. Thanks are extended to Steve Wagstaff and Russell Spangler for providing unpublished sequences and to Tom Philbrick, Claude dePamphilis, Andrea Wolfe, and Alison Colwell for providing plant material or DNA and for sharing insights and unpublished results.

[2] Department of E.P.O. Biology, University of Colorado, Boulder, Colorado 80309, U.S.A.

[3] Section of Molecular and Cellular Biology, University of California, Davis, California 95616, U.S.A.

and Selaginaceae relative to the "core" Scrophulariaceae, included them in the Scrophulariaceae. However, since Hallier (1903), classifications (e.g., Cronquist, 1981; Takhtajan, 1987; Thorne, 1992) have followed Wettstein (1895) in maintaining as distinct families the Lentibulariaceae, Orobanchaceae (except Takhtajan, 1987; Thorne, 1992), and Plantaginaceae. Infrafamilial classifications have recognized three subfamilies, Pseudosolaneae (primarily tribe Verbasceae), Antirrhinoideae, and Rhinanthoideae (e.g., Wettstein, 1895), or two subfamilies with Pseudosolaneae subsumed by Antirrhinoideae (e.g., Thieret, 1967), or, when Orobanchaceae is included, it is maintained as a third subfamily (Takhtajan, 1987; Thorne, 1992).

The placement of several genera putatively related to Scrophulariaceae, including a few monogeneric families (e.g., Callitrichaceae, Hippuridaceae, Hydrostachyaceae), has been disputed in angiosperm classifications. *Paulownia*, a large woody tree similar in habit, as well as flower, inflorescence, and leaf morphology, to *Catalpa* (Bignoniaceae), was placed in the Scrophulariaceae by Bentham and Hooker (Bentham, 1876) and Wettstein (1895), but was placed in the Bignoniaceae by Takhtajan (1980) and Cronquist (1981). However, Armstrong (1985) marshaled the evidence in favor of its placement in the Scrophulariaceae, and subsequent classifications have placed it there (e.g., Takhtajan, 1987; Thorne, 1992). *Schlegelia* and the related genera *Gibsoniothamnus* and *Synapsis* similarly have been bounced back and forth between the same two families, while not being considered closely related to *Paulownia*. Bentham and Hooker included *Schlegelia* in the Bignoniaceae, and this placement has been followed in the monograph of the family by Gentry (1980) and in Cronquist (1981) and Takhtajan (1987). Armstrong (1985) also considered *Schlegelia* and related genera and found that they did not fit with the Bignoniaceae (e.g., they have seeds with endosperm) and fit better within the Scrophulariaceae, although they are distinctive within that family. Thorne (1992) followed Armstrong (1985) by including the Schlegelieae in the Scrophulariaceae.

The Selagineae, a group of several genera and 100–200 species mostly native to South Africa and Madagascar, are distinguished by the presence of a single ovule in each of the two locules in the ovary (Thieret, 1967). This group was maintained as a distinct family by Bentham and Hooker (Bentham, 1876), but included in the Scrophulariaceae by Wettstein (1895) and most recent treatments (e.g., Takhtajan, 1987; Thorne, 1992). However, Cronquist (1981) placed the Selagineae in the Globulariaceae. The Globulariaceae are a small group of two genera and approximately 30 species (Thorne, 1992) that traditionally have been considered a distinct family, but also have been described as a tribe within the Scrophulariaceae (Barringer, 1993).

Aquatic angiosperms traditionally have been difficult to classify, and several aquatics have been associated with the Scrophulariales and related orders. The Lentibulariaceae are aquatic or semiaquatic and insectivorous, but retain showy insect-pollinated flowers that suggest a relationship with the Scrophulariaceae and have been classified close to that family or included within it (e.g., Hallier, 1903). However, the Callitrichaceae, Hippuridaceae, and Hydrostachyaceae, each a monogeneric family, all have very reduced floral morphology associated with either wind pollination (*Hippuris*, *Hydrostachys*) or water pollination (*Callitriche*), and their taxonomic placement has varied dramatically. Based primarily on embryological characters (e.g., the presence of unitegmic and tenuinucellate ovules) and evidence from floral development, these taxa have been placed in the Asteridae near the Scrophulariales or Lamiales (evidence reviewed in Wagenitz, 1992). Cronquist (1981) placed the three families together in the order Callitrichales, but this treatment has not been accepted widely (Wagenitz, 1992).

Root parasitism is a well-developed habit in the Scrophulariaceae, and the existence of green (photosynthetic) hemiparasites, such as *Orthocarpus*, *Pedicularis*, and *Striga*, is presumed to indicate a link to the achlorophyllous holoparasites of the Orobanchaceae. The dramatically reduced chloroplast genome in the Orobanchaceae (dePamphilis & Palmer, 1990), involving the loss of many genes involved with photosynthesis and other functions (Wolfe et al., 1992), has precluded the inclusion of representatives of the group in chloroplast DNA (cpDNA) studies involving either restriction site mapping of the entire genome or sequencing of *rbc*L. However, evidence from the cpDNA inverted repeat region (C. dePamphilis, pers. comm.) and DNA sequencing of the nuclear 18S rDNA (A. Colwell, pers. comm.) provide support linking the Orobanchaceae to the hemiparasites of the Scrophulariaceae.

A picture of phylogenetic relationships in the Asteridae has begun to emerge from studies of cpDNA, in which a monophyletic group, including the Scrophulariaceae and related families and provisionally designated the Lamiales s.l., has been identified (Downie & Palmer, 1992; Olmstead et al., 1992, 1993a). Initial studies of the chloroplast

gene *rbc*L were aimed at the identification and circumscription of major lineages within the Asteridae (Bremer et al., 1994; Olmstead et al., 1992, 1993a). Consequently, sampling of many families was inadequate to test hypotheses of monophyly at the family level. For previous cpDNA sequencing studies within the Lamiales s.l., sampling in the Scrophulariaceae has been limited to *Antirrhinum* and *Digitalis* (Olmstead et al., 1993a). Of the disputed genera and closely related families discussed above, only the Lentibulariaceae (Chase et al., 1993; Olmstead et al., 1993a) and Callitrichaceae (Olmstead et al., 1992, 1993a) have been sampled previously. Based on these analyses, the Lentibulariaceae were not closely related to the Scrophulariaceae, whereas *Callitriche*, usually placed more closely to the Labiatae (e.g., Wagenitz, 1992), forms a monophyletic group with *Antirrhinum* and *Digitalis* (Olmstead et al., 1993a).

In addition to the broad studies that have identified the Lamiales s.l. clade, other studies have addressed particular families within the Lamiales s.l. using cpDNA sequence data. Scotland et al. (1995), using *rbc*L and *ndh*F sequences, have identified a monophyletic Acanthaceae, which includes the often segregated Thunbergiaceae and Nelsoniaceae. Wagstaff & Olmstead (unpublished), using *rbc*L sequences, have confirmed a new division between the Labiatae and Verbenaceae, following Cantino's (1992) morphological cladistic study of the Labiatae. A monophyletic Labiatae has been established (Cantino et al., 1992), which includes much of the former Verbenaceae, leaving a reduced and apparently monophyletic Verbenaceae comprising only the former subfamily Verbenoideae. Also, a monophyletic Bignoniaceae has been circumscribed on the basis of *rbc*L and *ndh*F sequences (R. Spangler & R. Olmstead, unpublished), and similar work is under way on the Gesneriaceae (J. Smith, pers. comm.).

The goals of this study are to explore the possibility that the Scrophulariaceae (e.g., sensu Thorne, 1992) are not a monophyletic group and to examine the placement of several of the disputed genera and small families putatively allied with the Scrophulariaceae. The goal is not to establish an infrafamilial classification, because substantial doubt exists regarding whether the group is monophyletic in the first place. Attempts to discover an infrafamilial phylogeny using cpDNA inverted repeat restriction site mapping have met with frustration, for example, due to the lack of evidence for a monophyletic Scrophulariaceae (C. dePamphilis, pers. comm.). However, these efforts have discovered many groups of related genera, which ultimately may bear recognition at some level in a revised classification of the Lamiales s.l.

Sequencing the chloroplast gene *rbc*L has proven to be a useful approach for identifying phylogenetic relationships among higher-order groups in the angiosperms (e.g., *Annals of the Missouri Botanical Garden* Vol. 80(3), 1993) and has been particularly helpful in identifying major lineages within the Asteridae (Olmstead et al., 1993a). However, the highly conserved nature of the gene results in relatively few informative phylogenetic characters at the infrafamilial level. Sequencing studies focusing on individual families, with sufficient sampling to examine infrafamilial relationships (Conti et al., 1993; Doebley et al., 1990; Kim et al., 1992; Olmstead & Sweere, 1994; Soltis et al., 1993), have made clear the need to sample larger amounts of sequence to acquire sufficient numbers of characters to gain resolution and to assure greater accuracy in phylogenetic reconstructions. For this study, which crosses family lines to examine infra- as well as interfamilial relationships, a longer and more rapidly evolving cpDNA gene, *ndh*F (Olmstead & Sweere, 1994), was chosen to complement the substantial data set of *rbc*L gene sequences already available for representatives of the Lamiales s.l. By sequencing both genes for all taxa included in this analysis, nearly 3500 base pairs (bp) of cpDNA are available for phylogenetic inference, relative to the approximately 1400 bp compared in studies of *rbc*L alone.

The gene *ndh*F in tobacco encodes a protein of 740 amino acids, which is suggested to be a subunit of an NADH dehydrogenase enzyme (undiscovered at present) in the chloroplast (Sugiura, 1992). It is located in the small single-copy (SSC) region of the chloroplast genome close to the junction with the inverted repeat (IR). The termination codon of *ndh*F is located 43 bp from the IR/SSC boundary in tobacco, and transcription proceeds toward the inverted repeat. A comparison of rice and tobacco sequences suggests that *ndh*F has a nucleotide substitution rate approximately two times higher than that of *rbc*L (Sugiura, 1989). In combination with the fact that *ndh*F is approximately 50% longer than *rbc*L, the two-fold higher substitution suggests a three-fold increase in cladistic characters. In a previous study of the Acanthaceae (Scotland et al., 1995), almost exactly three times the number of characters were derived from *ndh*F relative to *rbc*L (421 vs. 136, note that taxon sampling was slightly different for the two genes); however, in a study of the Solanaceae (Olmstead & Sweere, 1994) *ndh*F provided only 60% more characters than *rbc*L (100 vs. 63).

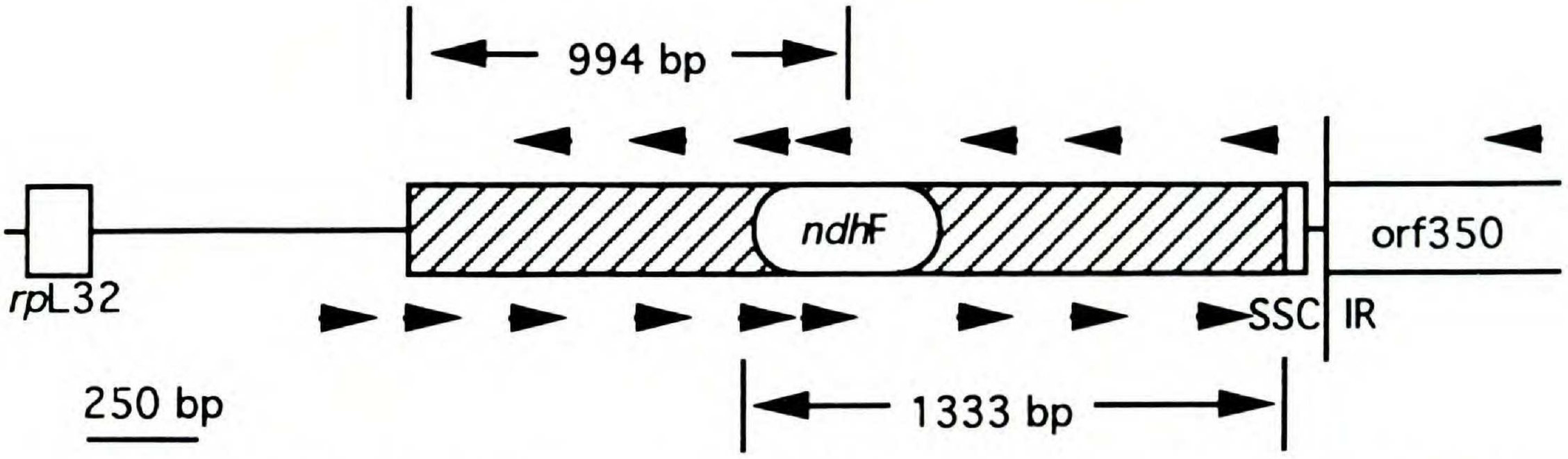


FIGURE 1. Map of *ndh*F and adjoining regions of cpDNA in tobacco. Boxes indicate reading frames; lines connecting boxes indicate noncoding DNA. Vertical bar at left end of orf350 indicates junction between inverted repeat (IR) and small single-copy region (SSC). Arrows indicate location of PCR and sequencing primers (primers located outside the coding region for *ndh*F supplied by R. Jansen, University of Texas). Cross-hatching in *ndh*F indicates portion sequenced. Overlapping regions marked above and below the map indicate the two fragments in which the gene was amplified.

MATERIALS AND METHODS

This study includes 29 representative species of the Lamiales s.l. and three outgroup species representing the most closely related orders; species names, voucher information, and GenBank accession numbers or prior publication references are given in Table 1. Eight tribes and both subfamilies Antirrhinoideae and Rhinanthoideae are represented (following Thieret, 1967), including six species classified as Scrophulariaceae by the most conservative treatments of the family: *Antirrhinum* (subfam. Antirrhinoideae, tribe Antirrhineae), *Celsia* (Antirrhinoideae, Verbasceae), *Scrophularia* (Antirrhinoideae, Scrophularieae), *Verbascum* (Antirrhinoideae, Verbasceae), *Digitalis* (Rhinanthoideae, Digitaleae), and *Veronica* (Rhinanthoideae, Veroniceae). Also included are disputed members *Paulownia* (Antirrhinoideae, Paulownieae), *Selago* (Antirrhinoideae, Selagineae), and *Schlegelia* (Antirrhinoideae, Schlegelieae), and representatives of three small putatively related families, Callitrichaceae, Hippuridaceae, and Plantaginaceae. Three additional species representing the Lentibulariaceae were included in an analysis of *rbc*L sequences with the 32 other species (results not shown). DNA sequences for the two genes were determined from the same plant accession for nearly all taxa. Exceptions include *Nicotiana* and *Barleria*, for which different accessions of the same species were used to determine the two gene sequences; *Thunbergia*, for which two different species were used, *T. usamberica* for *rbc*L and *T. alata* for *ndh*F; and *Verbena*, for which two different species were used, *V. bonariensis* for *rbc*L and *V. bracteata* for *ndh*F. Of the 64 sequences included in this study, 43 are not previously reported. It was determined during the course of this study that the previously published *rbc*L sequence of *Buddleja*, and the DNA accession from which it was derived, was mistakenly that of *Paulownia tomentosa* (cf. Bremer et al., 1994; Chase et al., 1993; Olmstead et al., 1993a; Scotland et al., 1995); the correct sequence for *Buddleja* is reported here.

Procedures for determining DNA sequences were as described in Olmstead & Sweere (1994). PCR amplification primers for *rbc*L are those described in Olmstead et al. (1992), with internal sequencing primers provided by G. Zurawski (DNAX Research Institute, Palo Alto, California). PCR amplification primers and internal sequencing primers for *ndh*F are described in Olmstead & Sweere (1994), with the exception that primers #1 and #2110R (forming the 5′ and 3′ ends of the amplified regions, respectively) were redesigned to match more exactly the homologous sequences for the Lamiales and were used to determine some sequences (Fig. 1). Redesigning the primers was made possible by sequencing through the ends of the gene in several species using primers located in flanking regions (primers kindly provided by R. Jansen, University of Texas). The new primer sequences are as follows (for others, see Olmstead & Sweere, 1994): #1—ATGGAACAGACATATCAATATG(C/G)GTGG and #2112R—CCC(C/T)A(C/G)ATATTTGATACCTTC(G/T)CC.

A total of 2135 bp of the *ndh*F gene was amplified in tobacco (somewhat more or less in other taxa depending on insertions and deletions) in two overlapping fragments of 994 bp and 1333 bp at the 5′ and 3′ ends of the gene respectively (Fig. 1). Approximately 97 bp at the 3′ end of the gene was not included. Preparation of DNA for sequenc-

TABLE 1. Sources of plant material, previously published sequences, and GenBank accession numbers for taxa included in this study (arrangement follows Thorne, 1992).

Family	Species	DNA Source/voucher[1]	*rbc*L	*ndh*F
Solanales				
Solanaceae	*Nicotiana tabacum* L.	Olmstead & Palmer, 1992	Lin et al., 1986	Olmstead et al., 1993b
Gentianales				
Gentianaceae	*Gentiana procera* Holm		Olmstead et al., 1993a	#L36400
Boraginales				
Boraginaceae	*Borago officinalis* L.		Olmstead et al., 1992	#L36393
Lamiales s.l.				
Acanthaceae	*Barleria prionitis* L.		Chase et al., 1993	Scotland et al., 1995
	Thunbergia usamberica Lindau		Chase et al., 1993	
	Thunbergia alata Bojer			Scotland et al., 1995
Bignoniaceae	*Catalpa* sp.		Olmstead et al., 1992	#L36397
	Martinella obovata (HBK) Bureau & K. Schum.	*Gentry 50277*	#L36444	#L36402
	Tabebuia heterophylla (A. de Candolle) Britton	Gentry (collection number unknown)	#L36451	#L36416
Buddlejaceae	*Buddleja davidii* Franch.		Olmstead et al., 1993a	#L36394
	Nicodemia diversifolia Tenore		Olmstead et al., 1993a	#L36405
Callitrichaceae	*Callitriche hermaphroditica* L.	*Philbrick 3022*	#L36441	#L36396
Gesneriaceae	*Nematanthus hirsutus* (Mart) Wiehler	Marie Selby B.G.	#L36446	#L36404
	Streptocarpus holstii Engl.		Olmstead et al., 1993a	#L36415
Hippuridaceae	*Hippuris vulgaris* L.	*Olmstead & Reeves 92-127*	#L36443	#L36401
Labiatae	*Ajuga reptans* L.		Wagstaff & Olmstead, unpublished	#L36391

TABLE 1. Continued.

Family	Species	DNA Source/voucher[1]	*rbc*L	*ndh*F
	Callicarpa dichotoma Koch		Olmstead et al., 1993a	#L36395
	Physostegia virginiana Benth.		Olmstead et al., 1993a	#L36407
Myoporaceae	*Myoporum mauritianum* A. de Candolle	*Olmstead 92-299*	#L36445	#L36403
Pedaliaceae	*Sesamum indicum* L.		Olmstead et al., 1993a	#L36413
Plantaginaceae	*Plantago lanceolata* L.	*Olmstead s.n.*	#L36454	#L36408
Scrophulariaceae	*Antirrhinum majus* L.		Olmstead et al., 1992	#L36392
	Celsia arcturus Jacq.	de Pamphilis (collection number unknown)	#L36442	#L36398
	Digitalis grandiflora Mill.		Olmstead et al,. 1993a	#L36399
	Paulownia tomentosa Steud.	*Olmstead 88-008*	#L36447	#L36406
	Selago thunbergii Choisy	*dePamphilis 90-21*	#L36450	#L36412
	Schlegelia parviflora (Oerst.) Monachino	*Gentry 14221*	#L36448	#L36410
	Scrophularia sp.	dePamphilis (collection number unknown)	#L36449	#L36411
	Verbascum thapsus L.	no voucher	#L36452	#L36417
	Veronica catenata Pennell	*Olmstead 92-144*	#L36453	#L36419
Verbenaceae	*Rhaphithamnus spinosus* (A. L. Juss.) Moldenke		Wagstaff & Olmstead, unpublished	#L36409
	Stachytarpheta dichotoma (Ruiz & Pav.) Vahl.		Wagstaff & Olmstead, unpublished	#L36414
	Verbena bonariensis L.		Olmstead et al., 1993a	
	Verbena bracteata Lagasca & Rodriguez	*Olmstead 92-131*		#L36418

[1] For voucher information on previously published sequences, see reference cited with sequence. References are cited in this column only when voucher information is cited in a reference other than the one in which the sequence is cited.

TABLE 2. Gaps in the sequence alignment for *ndh*F. Polarity (insertion vs. deletion) is relative to the results of the cladistic analysis of the nucleotide data. Coordinates are relative to the sequence for *Nicotiana tabacum*.

Insertion/deletion	Size (bp)	Coordinates[1]	Taxa
Insertion	6	657/658	*Nematanthus*
Insertion	6	1425/1426	*Selago*
Insertion/deletion	9	1443/1444	*Nicotiana*
Insertion	3	1443/1444	*Streptocarpus*
Deletion	42	1447–1488	*Callitriche*
Insertion	6	1467/1468	*Streptocarpus*
Insertion	6	1470/1471	*Selago*
Insertion/deletion	9	1471–1479	*Borago*
Insertion/deletion	9	1474–1482	*Gentiana*
Deletion	9	1477–1485	*Physostegia*
Insertion	12	1481/1482[2]	*Stachytarpheta*
Deletion	3	1486–1488	*Digitalis, Plantago, Veronica*
Deletion	6	1489–1494	*Veronica*
Deletion	6	1498–1503	*Barleria*
Insertion/deletion	9	1498–1506	*Gentiana*
Deletion	3	1501–1503	*Ajuga*
Deletion	9	1525–1533	*Verbena*
Insertion	6	1539/1540	all Lamiales s.l.; deleted in *Ajuga*, *Nicodemia*
Deletion	3	1564–1566	*Streptocarpus*
Insertion/deletion	6	1572/1573	*Gentiana*
Insertion	9	1695/1696	*Nematanthus, Plantago*
Insertion/deletion	6	1701/1702	*Gentiana, Plantago* (insertion)
Insertion/deletion	6	1702–1707	*Borago*
Deletion	6	1708–1713	*Barleria*
Insertion	6	1932/1933	*Streptocarpus*

[1] "/" indicates insertion between coordinates in *Nicotiana*; "–" range of nucleotides deleted relative to aligned *Nicotiana* sequence.

[2] Insertion not in frame with breaks between mRNA codons to enable identification of direct repeat as insertion.

ing proceeded by first amplifying each fragment with balanced primer concentrations to yield double-stranded DNA, then producing single-stranded DNA by a second round of PCR using a portion of the initial PCR product as template and a single primer (Kaltenboeck et al., 1992). Standard dideoxy sequencing was performed using Sequenase (U.S. Biochemical) and ^{32}P-labeled dATP. Each strand was sequenced for all taxa, requiring 12 sequencing reactions per *ndh*F sequence (Fig. 1).

Alignments of the sequences were determined by eye and, in the data set used for cladistic analysis, gaps in one species relative to any other were scored as missing data. The first 26 bp of the coding sequence of *rbc*L correspond to the 5′ PCR primer, and no gaps were identified through the first 1428 bp (the termination codon was beyond bp 1428 in all of the sequences), so 1402 bp of *rbc*L sequence were used for the phylogenetic analysis. Several gaps representing insertions, duplications, or deletions were observed in the *ndh*F sequences (Table 2). The first 23 bp of the coding sequence of *ndh*F correspond to the 5′ PCR primer, and the 3′ primer corresponds to bp 2110–2135 of the *Nicotiana tabacum* sequence. Inferred lengths of the *ndh*F genes sequenced for this study, including undetermined sequence downstream of the 3′ PCR primer, ranged in length from 2196 bp for *Callitriche heterophylla* to 2253 for *Nematanthus hirsutus*. The total aligned length including all gaps was 2167 bp.

The data were analyzed in three separate analyses: *rbc*L only, *ndh*F only, and both sequences combined. Parsimony analyses were conducted using PAUP version 3.1 (Swofford, 1993) with all changes weighted equally. For each analysis 100 random order entry replicate searches were conducted using the HEURISTIC search option with TBR branch swapping and with MULPARS "on" to save all equally most parsimonious trees. Results of each replicate were monitored to determine if multiple islands (Maddison, 1991) were discovered (as in Olmstead et al., 1993a). To assess the relative support for clades found in the combined analysis,

a bootstrap analysis was conducted with 100 replicates (Felsenstein, 1985). To examine possible circumscriptions of taxa that might represent a monophyletic Scrophulariaceae, four separate analyses were performed with each of the three data sets (each gene separately and the combined set) with various groups of taxa constrained to monophyly. The four constrained circumscriptions range from a strict definition of Scrophulariaceae, as found in many classifications, to a broad definition including many taxa of disputed family membership or belonging to small families inferred to be closely related to members of the Scrophulariaceae based on cpDNA studies. Following these constrained searches, a decay analysis (Donoghue et al., 1992) was performed on the combined data set, in which all trees were saved up to the length of the shortest tree in which a broadly defined monophyletic Scrophulariaceae was found. This last analysis was performed to determine how much structure remains in a tree based on these data if a monophyletic Scrophulariaceae is to be accepted. The data used in these analyses are available from the first author upon request, and sequences have been submitted to GenBank (Table 1).

Results

A total of 3569 bp of aligned sequence was used in this study, including 1402 bp (39%) of *rbc*L and 2167 aligned bp of *ndh*F. Of the 2167 bp of *ndh*F aligned sequence, 66 bp represent insertions or duplications unique to a single taxon or to two taxa and therefore are unable to provide any phylogenetic information, leaving 2101 bp of potentially useful aligned sequence. Of the 3569 total bp compared, 2196 bp (61.5%) were invariant, 640 bp (17.9%) were variable, but uninformative with respect to phylogeny, and 734 bp (20.6%) were phylogenetically informative. The *rbc*L sequences contributed 189 informative characters, or approximately 26% of the total while accounting for 40% of the useful aligned sequence, whereas *ndh*F sequences contributed 545 informative characters, or 74% of the total while accounting for only 60% of the useful aligned sequence. This 3 : 1 proportion in number of characters derived from *ndh*F relative to *rbc*L matches that found in the Acanthaceae (Scotland et al., 1995) and the prediction based on size and substitution rate of the two genes (Sugiura, 1989).

The presence of insertions and deletions (indels) in *ndh*F represents a significant difference relative to *rbc*L. In this study, 25 gaps in the sequence alignment are required to accommodate indels observed in the 32 sequences (Table 2). Of these, 21 are unique to an individual sequence (based on the results of the phylogenetic analysis, seven are insertions, eight are deletions, and six are among the outgroups and therefore unordered). Two gaps each apparently represent two independent insertions (inferred from the resulting tree based only on nucleotide substitutions) that are of the same size and in the same position in two unrelated sequences. In one case, both insertions represent direct repeats of a nine-bp adjacent sequence; in the other case, the six-bp insertions are not apparently related to adjacent sequences and share no nucleotides in common. The remaining two indels represent apparent synapomorphies. In one, a three-bp deletion is shared by *Digitalis*, *Veronica*, and *Plantago*, which together form a clade based on analysis of both of the gene sequences. In the other, a six-bp insertion forms an apparent synapomorphy for the entire Lamiales s.l., but gaps of the same size occur in the same position as the insertion in two ingroup sequences (*Nicodemia*, *Ajuga*) that are neither related to each other nor basal in the clade. Thus the 25 alignment gaps appear to represent 29 insertion or deletion events, of which two represent apparent synapomorphies of taxa in the current analysis. The phylogenetic significance of indels may be limited in this analysis, but their potential for phylogenetic inference at lower levels is great (see Discussion).

The parsimony analysis of *rbc*L sequences yielded eight equally most parsimonious trees (length = 851, consistency index, CI, = 0.426, excluding autapomorphies) occurring in two islands of six and two trees, respectively. The representatives of the Scrophulariaceae form two distinct clades (Figs. 2–4), designated "scroph I" (containing the type genus, *Scrophularia*, as well as *Celsia*, *Verbascum*, *Selago*, *Buddleja*, and *Nicodemia*) and "scroph II" (*Antirrhinum*, *Digitalis*, *Veronica*, *Plantago*, *Callitriche*, and *Hippuris*) on Figure 4. *Schlegelia* forms a clade with *Myoporum* in all *rbc*L trees and *Paulownia* forms a branch sister to the Bignoniaceae in the island of two trees or by itself near the base of the order in the island of six trees. Each island exhibits considerable resolution (trees not shown), but the strict consensus of all eight trees exhibits much less resolution (Fig. 2). The primary difference between the islands is in the placement of the Labiatae, which occur as sister group to a clade comprising the Acanthaceae and "scroph II" in island-2 (terminology of Maddison, 1991) and as a member of a large clade along with the Verbenaceae, Gesneriaceae, and "scroph I" in island-6. These results are congruent with those

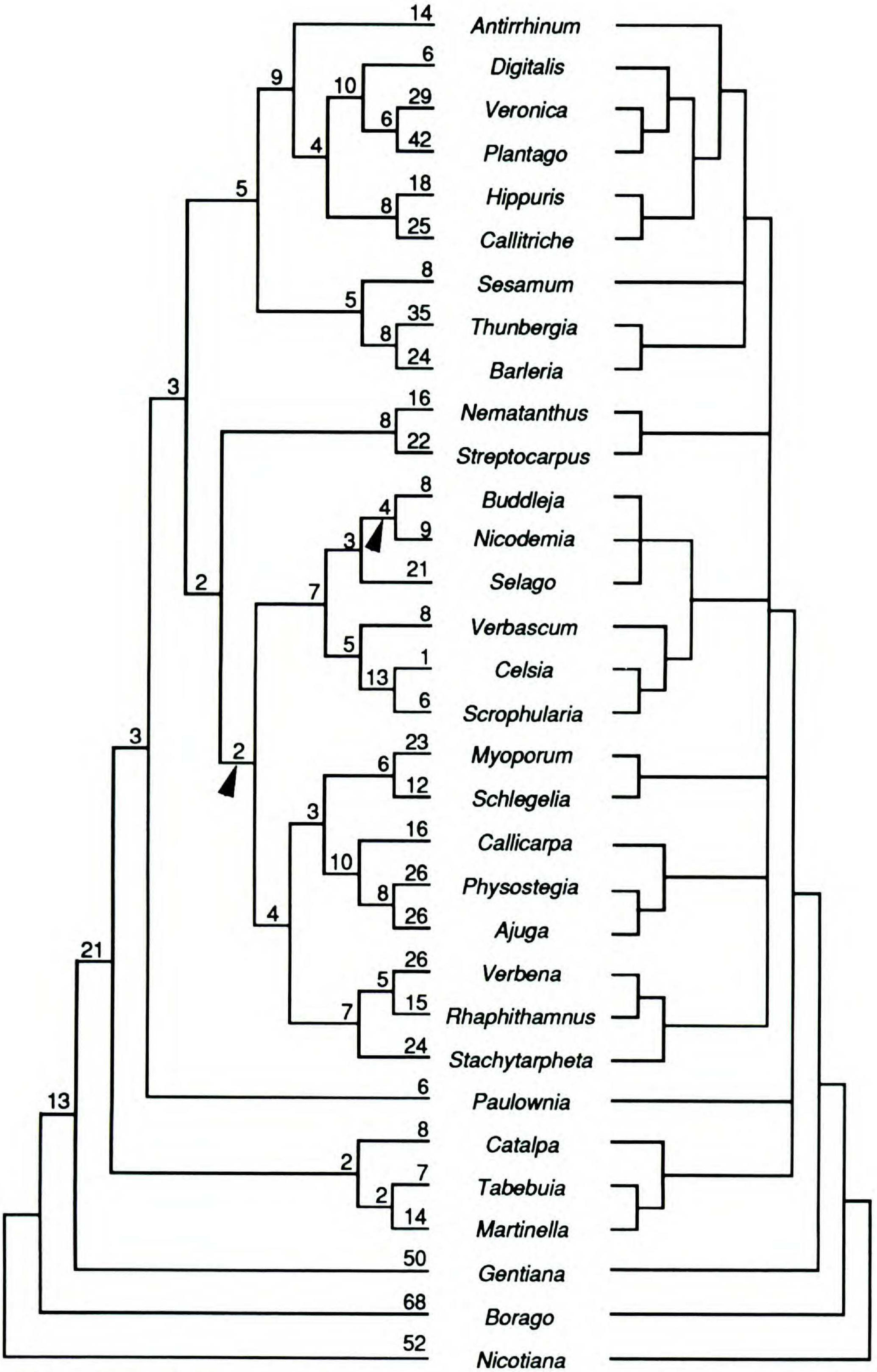


FIGURE 2. Results of parsimony analysis of *rbc*L sequences. On the left is one of six shortest trees in island-6 (length = 853) with branch lengths indicated. Arrows indicate internodes that collapse in the strict consensus of the six trees in island-6. On the right is a strict consensus of the eight shortest trees found, which includes the six trees in island-6 and the two trees in island-2.

obtained in previous analyses of *rbc*L sequences (e.g., Olmstead et al., 1993a).

The constrained search with a strict definition of Scrophulariaceae monophyly (*Antirrhinum*, *Celsia*, *Digitalis*, *Scrophularia*, *Selago*, *Verbascum*, *Veronica*) yielded 37 trees 24 steps longer than maximum parsimony (MP). A more relaxed constraint including *Plantago*, *Callitriche*, and *Hippuris* yielded 26 trees 9 steps longer than MP. Adding additional taxa, suggested by the maximum parsimony results to be potentially part of a monophyletic Scrophulariaceae clade, gave results as follows: (1) adding *Buddleja* and *Nicodemia*, 4 trees—3 steps; (2) adding *Buddleja*, *Nicodemia*, and *Myoporum*, 238 trees—8 steps. An additional analysis was performed, in which the *rbc*L sequences for *Byblis*, *Pinguicula*, and *Utricularia* (representing Lentibulariaceae) were retrieved from GenBank and added to the 32 sequences used elsewhere in this study. The results (not shown) indicate that the Lentibulariaceae represent a distinct lineage, not related to either of the two scroph clades identified in the analyses described above, confirming previously published results from a more limited sampling (Olmstead et al., 1993a).

The analysis of *ndh*F sequences yielded two trees (length = 2683, CI = 0.455, excluding autapomorphies). The representatives of the Scrophulariaceae form the same two distinct groups in the strict consensus tree (Fig. 3) as were found in the analysis of *rbc*L sequences. The closest groups to "scroph I" in order of increasing distance are Myoporaceae, Bignoniaceae, Acanthaceae, Verbenaceae, and Labiatae. "Scroph II" is sister group to the group just described. *Schlegelia* forms a clade with the Verbenaceae and *Paulownia* is sister group to the Labiatae. The constrained search with a strict definition of Scrophulariaceae yielded 4 trees 74 steps longer than MP. The more relaxed constraint (including *Plantago*, *Callitriche*, and *Hippuris*) yielded 10 trees 23 steps longer than MP. Adding additional taxa gave results as follows: (1) adding *Buddleja* and *Nicodemia*, 8 trees—8 steps; (2) adding *Buddleja*, *Nicodemia*, and *Myoporum*, 18 trees—4 steps.

The analysis of the combined *rbc*L and *ndh*F sequences yielded two trees (length = 3555, CI = 0.445 excluding autapomorphies). The same two distinct clades of Scrophulariaceae were obtained, and the branching order in the strict consensus tree (Fig. 4) is identical with respect to family-level groups to the results of the *ndh*F sequences. Results of the bootstrap analysis are shown in Figure 4. The constrained search with a strict definition of Scrophulariaceae yielded 3 trees 93 steps longer than MP. The more relaxed constraint (including *Plantago*, *Callitriche*, and *Hippuris*) yielded 5 trees 27 steps longer than MP. Adding additional taxa gave results as follows: (1) adding *Buddleja* and *Nicodemia*, 5 trees—9 steps; (2) adding *Buddleja*, *Nicodemia*, and *Myoporum*, 19 trees—4 steps. A decay analysis including all trees up to four steps longer than MP was performed to examine the degree of resolution remaining in the data, if a broadly defined Scrophulariaceae were to be accepted. In the strict consensus of the 1107 trees obtained, both "scroph I" and "scroph II" remain, as well as clades representing all of the families for which more than one representative was included (Buddlejaceae—within "scroph I," Acanthaceae, Bignoniaceae, Gesneriaceae, Labiatae, Verbenaceae) and the Lamiales s.l. as a whole. No other resolution within the ingroup is evident.

Discussion

Improved methods for obtaining DNA sequences have resulted in many systematists turning to direct sequence comparisons for molecular phylogenetic studies, rather than the indirect assessment of DNA sequence variation obtained using restriction site analysis. The initial widespread use of *rbc*L sequences was a product of two important factors: (1) the size and variability of the gene are appropriate for many systematic problems at the level of family and above and (2) primers for PCR amplification and sequencing were readily available to the plant systematics community. However, many of the problems of interest to plant systematists are at a level of phylogenetic divergence below which *rbc*L can provide sufficient information, or else represent problems of resolution among closely spaced branch points for which having more cladistic characters is desirable. Efforts to respond to these needs have resulted in the development of several alternative genes or regions of DNA for molecular phylogenetics (e.g., Baldwin, 1992; Johnson & Soltis, 1995, this issue; Olmstead & Sweere, 1994; Steele & Vilgalys, 1994), most of which have average substitution rates that are higher than that for *rbc*L (see Hoot et al., 1995, this issue, for an exception). For many studies at the infra- or interfamilial level, *rbc*L may still play a useful role in combination with another gene to increase the number of characters available (Olmstead & Sweere, 1994; Hoot et al., 1995, this issue; Johnson & Soltis, 1995; Soltis et al., 1993). In this study, the numbers of informative characters obtained from two genes, *rbc*L and *ndh*F, expressed in terms of characters per base pair of

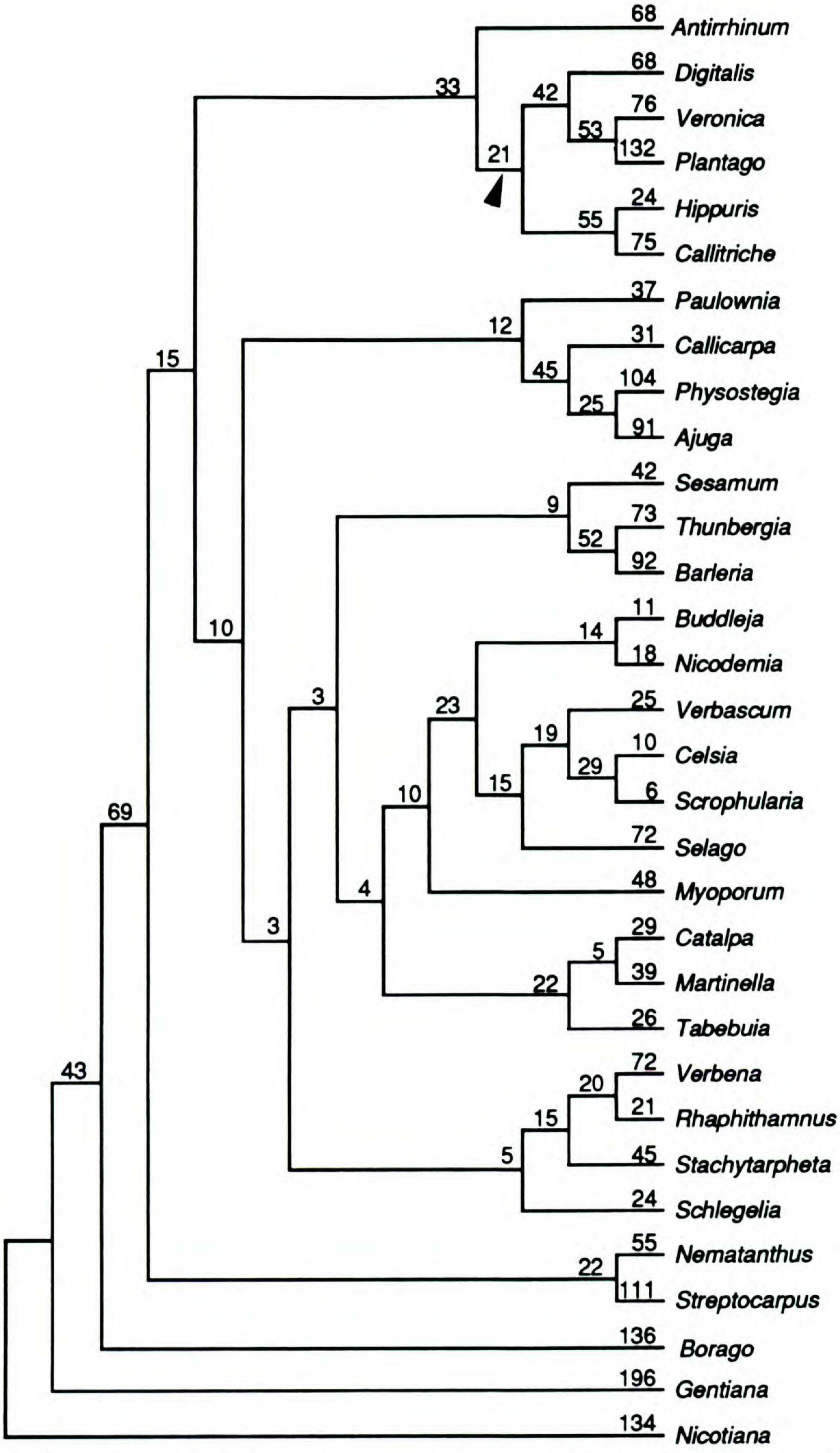


FIGURE 3. One of the two shortest trees resulting from parsimony analysis of *ndh*F sequences (length = 2683) with branch lengths indicated. Arrow indicates internode that collapses in strict consensus of the two trees.

useful aligned sequence, are 0.13 and 0.26 respectively, and the amount of sequence obtained was 50% greater for *ndh*F than for *rbc*L. Thus *ndh*F yielded three times as many characters as did *rbc*L. The problem of lineage identification (e.g., family-level clades) in the Lamiales s.l. and the relationship among them has been identified as one for which more data than are available from *rbc*L may be needed for resolution (Olmstead et al., 1993a).

*ndh*F sequences afford more than simply an additional source of nucleotide substitutions for phylogenetic inference. Indels are a regular, if infrequent, occurrence in the middle of the coding sequence of *ndh*F. This is unlike *rbc*L, for which insertions and deletions are unknown, except at the very end of the coding sequence where it is often difficult to discriminate insertions from point mutations in the termination codon resulting in an extension of the 3′ end. This feature of the gene may yield powerful phylogenetic characters that may be useful in many studies of infrafamilial phylogeny as has been demonstrated for *mat*K (Johnson & Soltis, 1995).

The alignment of protein-coding sequences containing indels is simplified, relative to noncoding DNA or rRNA coding genes, by the constraint of the genetic code for protein-coding genes. Alignments "by eye" are easier by virtue of including three nucleotides in the alignment frame. A group of relatively closely related species is unlikely to be so divergent that alignment matches will be obscured. In contrast, in noncoding or rRNA sequences indels may be of any length, and the problem of aligning one-base or two-base gaps is much greater. In the event that divergence of protein coding DNA sequences is great enough that alignment "by eye" is difficult, aligning the translated amino acid sequences may often resolve the placement of gaps that are not apparent in the DNA sequences.

The 25 alignment gaps in the set of *ndh*F sequences used in this study represent 29 evolutionary events as compared to the 2683 nucleotide substitution events (as inferred from the results of the cladistic analysis of the nucleotide information). The overall consistency index for nucleotide substitutions (including autapomorphies) is 0.56, whereas the same estimate for indels is 0.86. In this analysis, most of the indels were uninformative; however, each of the autapomorphic indels may be informative at a lower taxonomic level, as was found in several instances with *mat*K in the Saxifragaceae s.s. and Polemoniaceae (Johnson & Soltis, 1995). In the Solanaceae (Olmstead & Sweere, 1994), two additional indels were documented, each of which was unique to one of the 17 species included in that study. However, the nine-bp indel unique to *Nicotiana* in this study is common to all members of the Solanaceae examined and found in none of the other sequences yet obtained in the Asteridae, including those for the sister family Convolvulaceae, thereby providing a synapomorphy for the Solanaceae. In the Acanthaceae (Scotland et al., 1995), 14 indels were identified, of which 8 were unique to a single sequence and 6 were phylogenetically informative. Of the six informative ones, only one was an evident case of parallelism, and one other required two events in the shortest tree but could be explained as a single event in a tree two steps longer.

The observation of parallel occurrence of these infrequent events is more understandable when one takes into consideration the nonrandom distribution of indels in the sequence (Table 2, Fig. 5). Nineteen gaps, representing 21 inferred events, occur in a 148-bp region (between positions 1425 and 1573 in *Nicotiana*), and another 4 gaps (6 events) occur in an 18-bp region (1695–1713 in *Nicotiana*). Only two gaps, each representing a unique event, occur outside these regions (insertions at position 657/658 and 1932/1933). The higher incidence of indels is associated with higher substitution rates in the 3′ half of the *ndh*F sequence (Olmstead & Sweere, 1994), making that portion of the gene, by itself, appropriate for some phylogenetic studies (Catalan & Olmstead, unpublished).

The results of three separate analyses are presented (Figs. 2–4): *rbc*L sequences, *ndh*F sequences, and a combined data set with both sequences. The case has been made for conducting separate analyses for data sets derived from different genes (e.g., Swofford, 1991) in the event that one sequence exhibits strong interactions among positions that may bias substitutions, thereby violating the assumption of nonindependence among characters. However, no evidence of such differences has been found in prior analyses of *rbc*L and *ndh*F sequences (Olmstead & Sweere, 1994; Scotland et al., 1995), and none is anticipated here. Performing the separate analysis does afford the opportunity to examine the efficacy of each gene for phylogeny reconstruction and to examine the effect of partitioning a data set to see whether strongly supported clades in one partition are equally supported in another (Olmstead & Sweere, 1994). The results of the analyses of the two individual data sets for this group of taxa (Figs. 2–3) are congruent with respect to finding the same clades that are the most well-supported clades in the combined analysis (Fig.

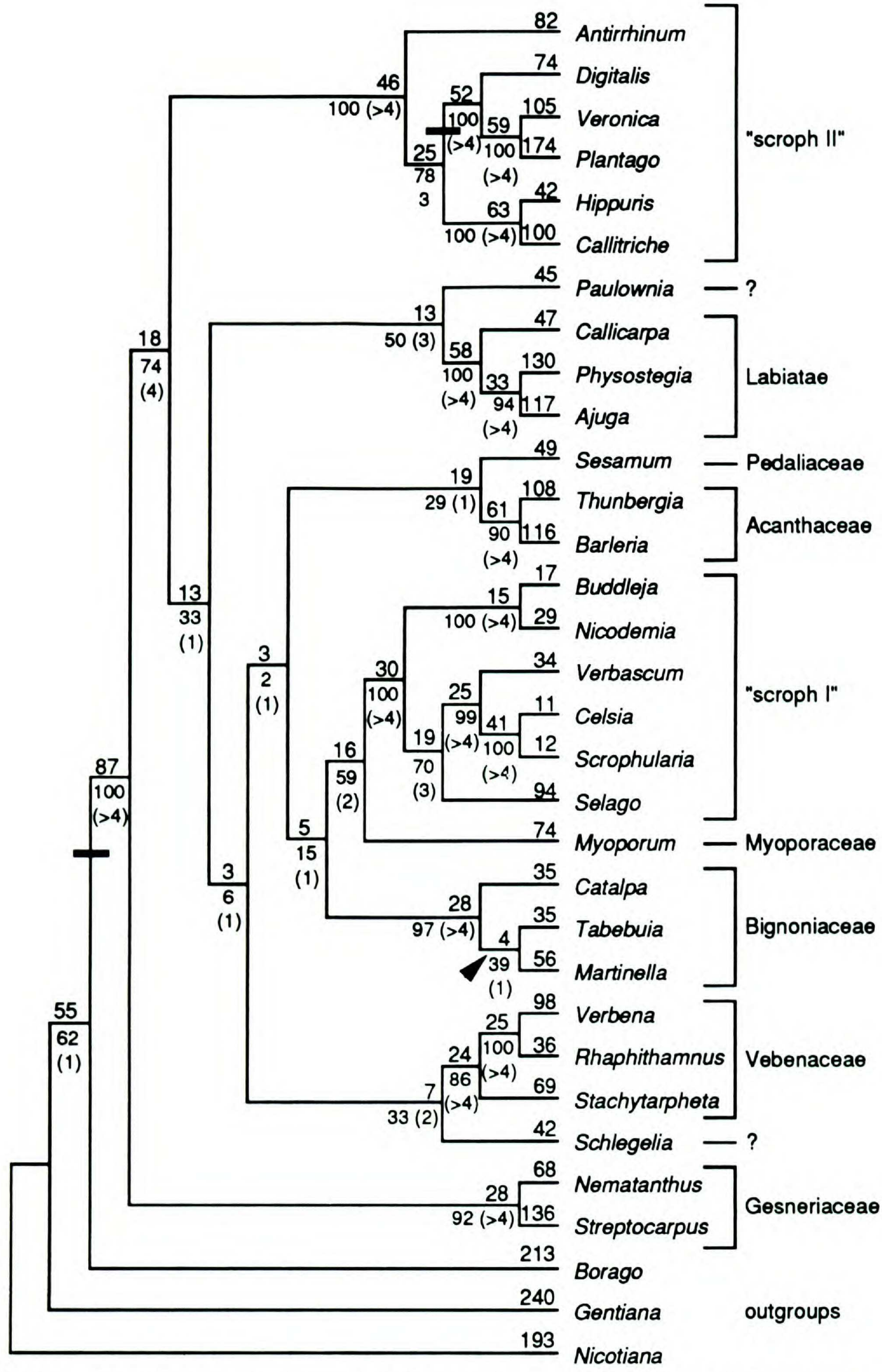


FIGURE 4. One of the two shortest trees resulting from parsimony analysis of *rbc*L and *ndh*F sequences combined (length = 3555) with branch lengths indicated above the internodes (the other tree is the same as Fig. 3). Numbers below the internodes indicate bootstrap and decay (in parentheses) values. Arrow indicates internode that collapses

4). The differences between the two results occur in areas where support is weak in both individual trees. Groups identified by *rbc*L, but not *ndh*F, all had five or fewer synapomorphies on the *rbc*L tree, and groups identified by *ndh*F, but not *rbc*L, all had 12 or fewer synapomorphies on the *ndh*F tree; groups in common in both results had much higher character support. The results of the combined analysis (Fig. 4) are nearly identical to the results of the *ndh*F analysis (Fig. 3), with one of the two most parsimonious trees found in each analysis being identical. However, *rbc*L informs the combined result in several ways: (1) *ndh*F does not fully resolve internal relationships in the "scroph II" clade, whereas *rbc*L and the combined analyses do. (2) *ndh*F and *rbc*L resolve the relationships among the three representatives of the Bignoniaceae differently; the combined analysis remains unresolved. (3) Most importantly, at almost every node where the combined tree agrees with the *ndh*F result, but not the *rbc*L result, *rbc*L provides additional characters in support of the conclusions, indicating that phylogenetic signal is present in the "*rbc*L partition" for some portions of the tree where very few characters are informative and may be outweighed by chance homoplasy in that relatively small partition of the data.

SYSTEMATIC CONCLUSIONS

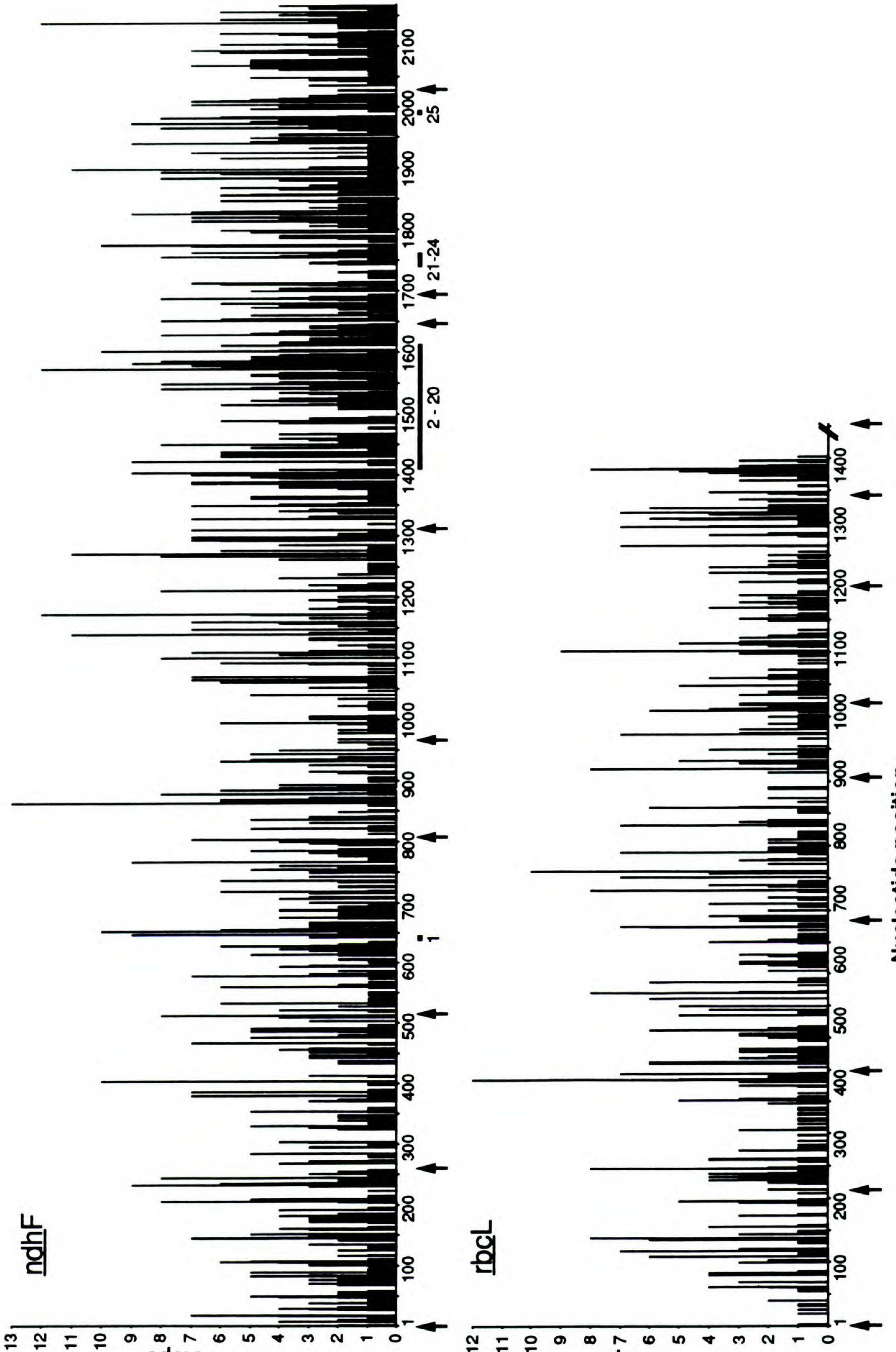
The primary systematic conclusion of this study is that the Scrophulariaceae, as circumscribed in most taxonomic treatments, are polyphyletic. The polyphyly of the Scrophulariaceae is complex in the following ways: (1) There exist two distinct clades in this study (and possibly more in nature) that contain elements of the traditional Scrophulariaceae (referred to as "scroph I" and "scroph II" in Fig. 4). (2) There are groups that are excluded from the family in traditional classifications that are clearly related to and derived from elements of the traditional Scrophulariaceae (e.g., Plantaginaceae, Callitrichaceae, Hippuridaceae, and apparently Orobanchaceae, as noted previously, which is not included in this study). (3) Some disputed genera, assigned to the Scrophulariaceae in some treatments (e.g., *Schlegelia* and *Paulownia* in Thorne, 1992) do not belong to either "scroph I" or "scroph II."

The results of cladistic analysis of data sets derived from both *rbc*L and *ndh*F identify most of the same terminal lineages within the Lamiales s.l. Represented in these lineages are the Acanthaceae, Labiatae, Verbenaceae, Bignoniaceae, Gesneriaceae, and two groups of taxa containing elements of the traditional Scrophulariaceae. The "scroph I" lineage consists, in this study, primarily of elements that are distinctive and often isolated within or near the traditional Scrophulariaceae: *Verbascum* and *Celsia* (Antirrhinoideae, Verbasceae), *Selago* (Antirrhinoideae, Selagineae), *Buddleja* and *Nicodemia* (Buddlejaceae), which are often placed close to the Scrophulariaceae, but not within it (e.g., Cronquist, 1981), and the type genus *Scrophularia* (Antirrhinoideae, Scrophularieae). The Verbasceae, which have nearly actinomorphic flowers, were previously classified as subfamily Pseudosolaneae (Wettstein, 1895) which suggested connections to the Solanaceae. In fact, Bentham and Hooker (Bentham, 1876) placed the solanaceous tribe Salpiglossideae, which have zygomorphic flowers, unlike most Solanaceae, in the Scrophulariaceae in recognition of the similarity. The Selagineae have been segregated from the Scrophulariaceae (e.g., Bentham & Hooker in Bentham, 1876) into their own family primarily on the basis of the reduction in ovule number to one per carpel (Thieret, 1967). Reduction in ovule number to one or two per carpel, at times with each carpel subdivided into two uniovulate mericarps, has occurred independently in the Lamiales s.l. on several occasions (Wagstaff & Olmstead, unpublished). The Buddlejaceae are provisionally included in the "scroph I" clade, rather than as sister group to it (as suggested by the *ndh*F and combined results), because the *rbc*L results include them within the otherwise completely scroph group and the bootstrap and decay values determined for the combined data set do not unequivocally reject that hypothesis. Recognition of the fact that the genus *Scrophularia* belongs in this group means that a phylogenetic classification would have to recognize "scroph I" or some portion of it as Scrophulariaceae, and "scroph II" would require a new family name.

The "scroph II" lineage consists of the traditional scrophs *Antirrhinum* (Antirrhinoideae, Antirrhineae), *Digitalis* (Rhinanthoideae, Digitaleae),

←

in strict consensus of the two trees. Dark bars indicate a deletion (within "scroph II") and an insertion (at the base of the Lamiales s.l.) in the *ndh*F sequences. Question marks (?) next to *Paulownia* and *Schlegelia* indicate that their classification into a higher order taxon remains uncertain.


ndhF
rbcL
Steps
Nucleotide position
1
2 - 20
21-24
25

and *Veronica* (Rhinanthoideae, Veroniceae), and representatives of the derived families Callitrichaceae, Hippuridaceae, and Plantaginaceae. Thieret (1967) suggested that the Scrophulariaceae be divided into two subfamilies with the Antirrhinoideae ancestral to the Rhinanthoideae. In this analysis, members of the Antirrhinoideae are included in both "scroph I" and "scroph II" lineages, making the subfamily polyphyletic, whereas the sampled members of the Rhinanthoideae form a monophyletic group with *Plantago*. The Rhinanthoideae contain all of the hemiparasitic Scrophulariaceae, from which the Orobanchaceae are evidently derived (C. dePamphilis, pers. comm.; A. Colwell, pers. comm.), thereby suggesting that *Orobanche* and relatives belong in this clade. The Plantaginaceae exhibit modifications to floral morphology associated with wind pollination, which have resulted in their segregation as a separate family, while retaining enough similarity to be recognized as close to or derived from the Scrophulariaceae. However, the aquatic families Callitrichaceae and Hippuridaceae exhibit extremely modified reproductive morphology, which has baffled previous efforts to classify them with their closest relatives in the Scrophulariaceae. The third aquatic family in Cronquist's (1981) Callitrichales, the Hydrostachyaceae, has been found to be unrelated to the Callitrichaceae and Hippuridaceae and is most closely related to the Hydrangeaceae in the Cornales on the basis of *rbc*L sequences (Hempel et al., in press).

The phylogenetic position and classification of *Schlegelia* and *Paulownia* remain unclear. *rbc*L and *ndh*F sequences each suggest a different placement for *Schlegelia* (near Myoporaceae—*rbc*L; near Verbenaceae—*ndh*F) and *Paulownia* (isolated near the base of the order or with the Bignoniaceae—*rbc*L; near Labiatae—*ndh*F). The combined analysis provides relatively weak evidence for the placement of *Schlegelia* near the Verbenaceae (33% bootstrap value) and *Paulownia* near the Labiatae (50% bootstrap value). The evidence does suggest that neither genus belongs with either of the lineages containing other members of the Scrophulariaceae. *Schlegelia*, and the related genera *Gibsoniothamnus* and *Synapsis*, may represent one of several small lineages in the Lamiales s.l. that have been "shoehorned" into families representing larger lineages in the past. The Lentibulariaceae appear to represent another small independent lineage (Olmstead et al., 1993a) that has been suggested incorrectly to be a derivative of the Scrophulariaceae (e.g., Cronquist, 1981). An expanded analysis of *rbc*L sequences, including three representatives of the Lentibulariaceae in addition to the 32 taxa in this study, indicates that the Lentibulariaceae are not related to either of the scroph lineages identified here (results not shown).

Given that the phylogenetic hypothesis derived from the parsimony analysis of the cpDNA sequences presented here may not be the true phylogeny of the group, it is instructive to examine prospective circumscriptions of groups that may represent a monophyletic Scrophulariaceae. This was done by imposing on the parsimony analysis constraints that require a group to be monophyletic, while finding the most parsimonious trees. Several prospective circumscriptions were examined, beginning with a very strict definition matching most recent classifications and expanding to a broad circumscription that includes all taxa belonging to the lineages "scroph I" and "scroph II" and *Myoporum*. Not surprisingly, parsimony analyses constrained to the narrowest circumscriptions resulted in trees that were much longer than maximum parsimony (e.g., 93 steps longer for the most conservative definition of Scrophulariaceae in the combined analysis of *rbc*L and *ndh*F sequences). However, a broadly circumscribed Scrophulariaceae, including the Plantaginaceae, Callitrichaceae, Hippuridaceae, Buddlejaceae, and *Myoporum* are monophyletic in trees four steps longer than maximum parsimony in the combined analysis.

If the results of a parsimony analysis were to be used as evidence for classification in which a group that requires trees four steps longer than maximum parsimony might be acceptable, then it is relevant to ask how many trees are found and how much resolution is retained when all trees up

←

FIGURE 5. Histograms representing the changes by nucleotide position inferred over the shortest estimated tree for the combined data (tree in Fig. 4) for the genes *ndh*F and *rbc*L. The nucleotide position numbers refer to the aligned sequences used in the phylogenetic analysis beginning with the first nucleotide after the 5′ PCR primer and ending with the last nucleotide before the 3′ PCR primer. Arrows indicate locations of the PCR and sequencing primers. The bars under the *ndh*F sequence indicate the locations of gaps in the sequences (numbers refer to Table 2).

to four steps longer than the shortest are examined. A decay analysis saving all trees up to four steps longer than maximum parsimony found 1107 trees. The consensus among those trees shows that all families for which more than one representative were included are supported and that both "scroph I" and "scroph II" lineages are supported, but no other resolution is possible within the Lamiales s.l. Whereas there is no assurance that the shortest tree is the true phylogeny for this group of taxa, the fact that the "scroph I" and "scroph II" lineages are obtained even in the consensus of the trees up to four steps longer suggests that their classification as two distinct groups comparable to other groups resolved at this level (i.e., family) would be a conservative move and would produce a more stable classification when further evidence is obtained.

The debate has subsided, but no consensus has been reached on the distinction between polyphyly and paraphyly (or even monophyly—see Cronquist, 1987). Even Hennig used different criteria for the terms, first defining polyphyly on the basis of shared, convergent characters and paraphyly on the basis of shared, ancestral characters (Hennig, 1966), then (Hennig, 1975) distinguishing the two by whether the common ancestral (stem) species of a group would be included (paraphyly) or not included (polyphyly) in that group. The point at which the debate impacts systematics and classification most strongly centers on whether a greater distinction needs to be made between monophyletic versus para- and polyphyletic groups (Donoghue & Cantino, 1988) or mono- and paraphyletic versus polyphyletic groups (Cronquist, 1987). This study serves a useful purpose by illuminating how difficult and pointless it is to try to distinguish between paraphyly and polyphyly as opposed to monophyly. Perhaps a group that is defined on the basis of plesiomorphic characters or by the inclusion of a stem species could be recognized and named Scrophulariaceae, but such a classification would serve poorly any evolutionary applications, because virtually all of the Lamiales s.l. could be included in an evolutionary lineage identified by those traits or derived from that stem species. The fact that progress is being made to identify and circumscribe monophyletic families successfully in the Lamiales s.l. argues against the continued recognition of a traditional Scrophulariaceae and for the recognition of family-level groups that were once part of the Scrophulariaceae. The circumscription and formal recognition of such groups awaits the completion of work in progress (C. dePamphilis, R. Olmstead & A. Wolfe, unpublished). It is hoped that the recognition of this division within the Scrophulariaceae based on cpDNA will encourage additional morphological and anatomical work aimed at identifying monophyletic groups containing elements of the traditional Scrophulariaceae.

Literature Cited

Armstrong, J. E. 1985. The delimitation of Bignoniaceae and Scrophulariaceae based on floral anatomy, and the placement of problem genera. Amer. J. Bot. 72: 755–766.

Baldwin, B. G. 1992. Phylogenetic utility of the transcribed spacers of nuclear ribosomal DNA in plants: An example from the Compositae. Molec. Phylogenetics Evol. 1: 1–16.

Barringer, K. 1993. Five new tribes in the Scrophulariaceae. Novon 3: 15–17.

Bentham, G. 1876. Scrophulariaceae. Pp. 913–980 *in* G. Bentham & J. D. Hooker (editors), Genera Plantarum, Vol. 2. Reeve, London.

Bremer, B., R. G. Olmstead, L. Struwe & J. A. Sweere. 1994. *rbc*L sequences support exclusion of *Retzia*, *Desfontainia*, and *Nicodemia* from the Gentianales. Pl. Syst. Evol. 190: 213–230.

Cantino, P. D. 1992. Evidence for a polyphyletic origin of the Labiatae. Ann. Missouri Bot. Gard. 79: 361–379.

———, R. M. Harley & S. J. Wagstaff. 1992. Genera of Labiatae: Status and classification. Pp. 511–522 *in* R. Harley (editor), Advances in Labiate Science. Royal Botanic Gardens, Kew.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H. G. Hills, Y.-L. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedren, B. S. Gaut, R. K. Jansen, K.-J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, Q.-Y. Xiang, G. M. Plunkett, P. S. Soltis, S. M. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, S. Graham, S. C. H. Barrett, S. Dayanandan & V. A. Albert. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Missouri Bot. Gard. 80: 528–580.

Conti, E., A. Fischbach & K. J. Sytsma. 1993. Tribal relationships in Onagraceae: Implications from *rbc*L sequence data. Ann. Missouri Bot. Gard. 80: 672–685.

Cronquist, A. 1981. An Integrated System of Classification of Flowering Plants. Columbia Univ. Press, New York.

———. 1987. A botanical critique of cladism. Bot. Rev. 53: 1–52.

dePamphilis, C. W. & J. D. Palmer. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348: 337–339.

Doebley, J., M. Durbin, E. M. Golenberg, M. T. Clegg & D. P. Ma. 1990. Evolutionary analysis of the large subunit of carboxylase (*rbc*L) nucleotide sequence among the grasses (Gramineae). Evolution 44: 1097–1108.

Donoghue, M. J. & P. D. Cantino. 1988. Paraphyly, ancestors, and the goals of taxonomy: A botanical defense of cladism. Bot. Rev. 54: 107–128.

———, R. G. Olmstead, J. F. Smith & J. D. Palmer. 1992. Phylogenetic relationships of Dipsacales based on *rbc*L sequences. Ann. Missouri Bot. Gard. 79: 333–345.

Downie, S. R. & J. D. Palmer. 1992. Restriction site mapping of the chloroplast DNA inverted repeat: A molecular phylogeny of the Asteridae. Ann. Missouri Bot. Gard. 79: 266–283.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.

Gentry, A. H. 1980. Bignoniaceae. Flora Neotropica, monograph number 25. New York Botanical Garden, New York.

Hallier, H. 1903. Ueber die Abgrenzung und Verwandtschaft der einzelnen Sippen bei den Scrophulariaceen. Bull. Herb. Boiss., II 3: 181–207.

Hempel, A. L., P. A. Reeves, R. G. Olmstead & R. K. Jansen. Implications of *rbc*L sequence data for higher order relationships of the Loasaceae and the anomalous aquatic plant *Hydrostachys* (Hydrostachyaceae). Pl. Syst. Evol., in press.

Hennig, W. 1966. Phylogenetic systematics. Univ. Illinois Press, Urbana.

———. 1975. "Cladistic analysis or cladistic classification?": A reply to Ernst Mayr. Syst. Zool. 24: 244–256.

Hoot, S. B., A. Culham & P. R. Crane. 1995. The utility of *atp*B gene sequences in resolving phylogenetic relationships: Comparison with *rbc*L and 18S ribosomal DNA sequences in the Lardizabalaceae. Ann. Missouri Bot. Gard. 82: 194–208.

Johnson, L. A. & D. E. Soltis. 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *mat*K sequences. Ann. Missouri Bot. Gard. 82: 149–175.

Kaltenboeck, B., J. W. Spatafora, X. Zhang, K. G. Kousoulas, M. Blackwell & J. Storz. 1992. Efficient production of single-stranded DNA as long as 2 kb for sequencing of PCR-amplified DNA. BioTechniques 12: 164–171.

Kim, K.-J., R. K. Jansen, R. S. Wallace, H. J. Michaels & J. D. Palmer. 1992. Phylogenetic implications of *rbc*L sequence variation in the Asteraceae. Ann. Missouri Bot. Gard. 79: 428–445.

Lin, C. M., Z. Q. Liu & S. D. Kung. 1986. *Nicotiana* chloroplast genome: X. Correlation between the DNA sequences and the isoelectric focusing patterns of the LS of Rubisco. Pl. Molec. Biol. 6: 81–87.

Maddison, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. Syst. Zool. 40: 315–328.

Olmstead, R. G. & J. D. Palmer. 1992. A chloroplast DNA phylogeny of the Solanaceae: Subfamilial relationships and character evolution. Ann. Missouri Bot. Gard. 79: 346–360.

——— & J. A. Sweere. 1994. Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the Solanaceae. Syst. Biol. 43: 467–481.

———, ———, & K. H. Wolfe. 1993b. Ninety extra nucleotides in *ndh*F gene of tobacco chloroplast DNA: A summary of revisions to the 1986 genome sequence. Pl. Molec. Biol. 22: 1191–1193.

———, B. Bremer, K. M. Scott & J. D. Palmer. 1993a. A parsimony analysis of the Asteridae sensu lato based on *rbc*L sequences. Ann. Missouri Bot. Gard. 80: 700–722.

———, H. J. Michaels, K. M. Scott & J. D. Palmer. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbc*L. Ann. Missouri Bot. Gard. 79: 249–265.

Scotland, R. W., J. S. Sweere, P. A. Reeves & R. G. Olmstead. 1995. Higher level systematics of Acanthaceae determined by chloroplast DNA sequences. Amer. J. Bot., in press.

Soltis, D. E., D. R. Morgan, A. Grable, P. S. Soltis & R. Kuzoff. 1993. Molecular systematics of Saxifragaceae sensu stricto. Amer. J. Bot. 80: 1056–1081.

Steele, K. P. & R. Vilgalys. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *mat*K. Syst. Bot. 19: 126–142.

Sugiura, M. 1989. The chloroplast chromosomes in land plants. Ann. Rev. Cell Biol. 5: 51–70.

———. 1992. The chloroplast genome. Pl. Molec. Biol. 19: 149–168.

Swofford, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? Pp. 295–333, *in* M. M. Miyamoto & J. Cracraft (editors), Phylogenetic Analysis of DNA Sequences. Oxford Univ. Press, New York.

———. 1993. PAUP: Phylogenetic Analysis Using Parsimony, vers. 3.1. Illinois Natural History Survey, Champaign, Illinois.

Takhtajan, A. 1980. Outline of the classification of flowering plants (Magnoliophyta). Bot. Rev. 46: 225–359.

———. 1987. Systema Magnoliophytorum. Nauka, Leningrad. [In Russian.]

Thieret, J. W. 1967. Supraspecific classification in the Scrophulariaceae: A review. Sida 3: 87–106.

Thorne, R. T. 1992. Classification and geography of the flowering plants. Bot. Rev. 58: 225–348.

Wagenitz, G. 1992. The Asteridae: Evolution of a concept and its present status. Ann. Missouri Bot. Gard. 79: 209–217.

Wettstein, R. von. 1895. Scrophulariaceae. Pp. 39–107 *in* A. Engler & K. Prantl (editors), Die Natürlichen Pflanzenfamilien. Vol. 4(3b). Wilhelm Engelmann, Leipzig.

Wolfe, K. H., C. W. Morden & J. D. Palmer. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc. Natl. Acad. Sci., U.S.A. 89: 10648–10652.